Knowledge Management and Taxonomies

Dr. Geoffrey P Malafsky, TechI^2 LLC

TechI2

Context: Who Are We?

- Technology Architect: Knowledge and Content Management Systems
  - Dept of Navy CIO: Taxonomy, systems, business processes, metrics
  - Lifelines Portal: redesign of 20 million hits/month public site to Oracle architecture with workflow + portal + taxonomy + search
- Independent Assessments and Planning
  - Balanced Scorecards: technology ↔ resources ↔ processes
  - External Independent Review: in-depth objective analysis of all good and bad aspects of system or plan
  - R&D Expert:
    - DARPA: LCCMD (MEMS), SUO-SAS (mobile peer-to-peer data fusion and dissemination); CPOF (KM and visualization)
    - ONR: Swampworks; CINC21

The Need

- Share and reuse knowledge
- Reduce information overload
- Minimize operation and maintenance costs
- Streamline business processes
- Enterprise system interoperability

The Challenge

- Enormous quantity of written, spoken, and visual information
- Confusion about what “knowledge” is
- Language complexity and dynamism
- Limited tool accuracy
- Multiple systems and legacy applications

An Approach

- Develop and implement the proper set of taxonomies for the organization, workflow, and subject matter
- Incorporate contextual meaning into system architecture with business logic

What is Knowledge?

- Context: what is it about?
- Confidence: is it right?
- Relationships: what does it have to do with that?
- Priorities: what is most important?
- Types
  - Explicit knowledge is codified and can be manipulated
  - Tacit knowledge is unspoken “know-how”

Knowledge Ontology:Bloom 1956

- Knowledge
  - Knowledge of specifics
    - knowledge of terminology
    - knowledge of specific facts
  - Knowledge of ways and means of dealing with specifics
    - knowledge of conventions
    - knowledge of trends and sequences
    - knowledge of classifications and categories

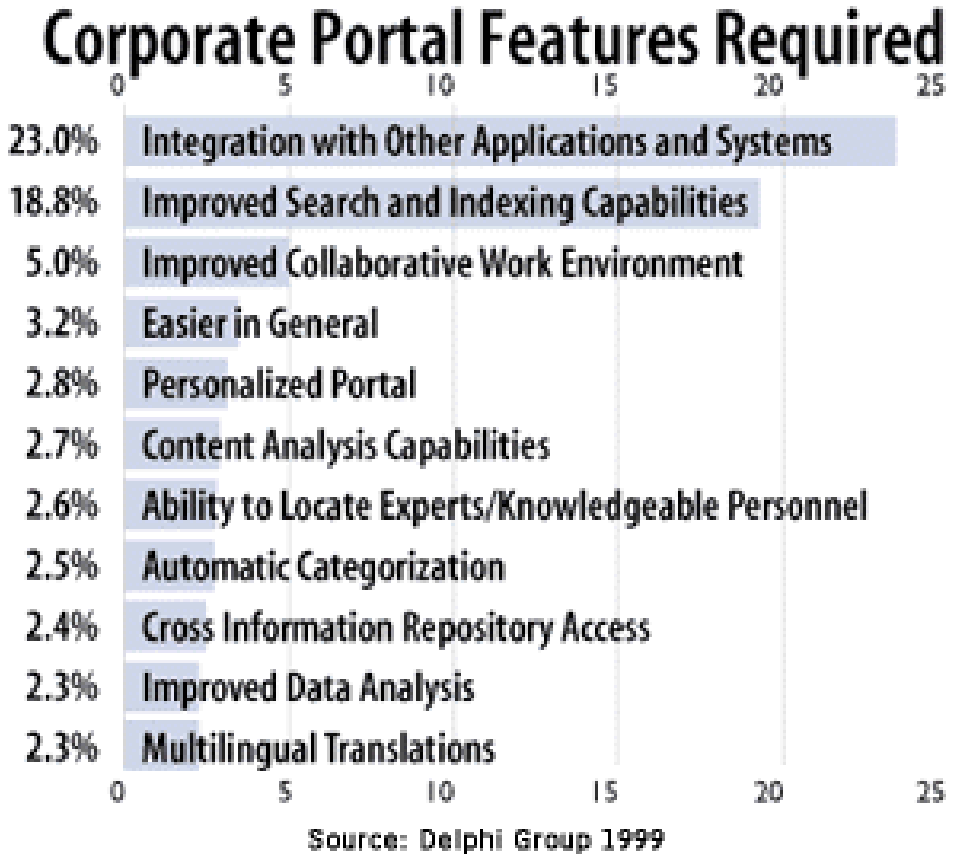
Knowledge is Personal

- Depends on prior knowledge, and task focus
- "Set the soldering iron to 200 degrees"
  - *information* from manual for general use
  - *knowledge* from expert for specific manufacturing process
- "10000 units shipped yesterday"
  - *data* for logistics
  - *information* for shipping manager
  - *knowledge* for competitor monitoring market share

Language Complexity

- People need the metamessages conveyed by human interactions
  - body language
  - clothing
  - dialect
  - intonation

Where is it?


Corporate Portal Features Required
0
5
10
15
20
25
23.0% Integration with Other Applications and Systems
18.8% Improved Search and Indexing Capabilities
5.0% Improved Collaborative Work Environment
3.2% Easier in General
2.8% Personalized Portal
2.7% Content Analysis Capabilities
2.6% Ability to Locate Experts/Knowledgeable Personnel
2.5% Automatic Categorization
2.4% Cross Information Repository Access
2.3% Improved Data Analysis
2.3% Multilingual Translations
0
5
10
15
20
25
Source: Delphi Group 1999

Semantic Web

- The vision of the *Semantic Web* aims at creating a Web where information can be "understood" by machines as well as humans
- The *Semantic Web* requires the emergence of a general purpose representation and markup-language to convey information about machine-accessible semantics
- Taxonomies can be expressed as abbreviated Ontologies for the *Semantic Web*

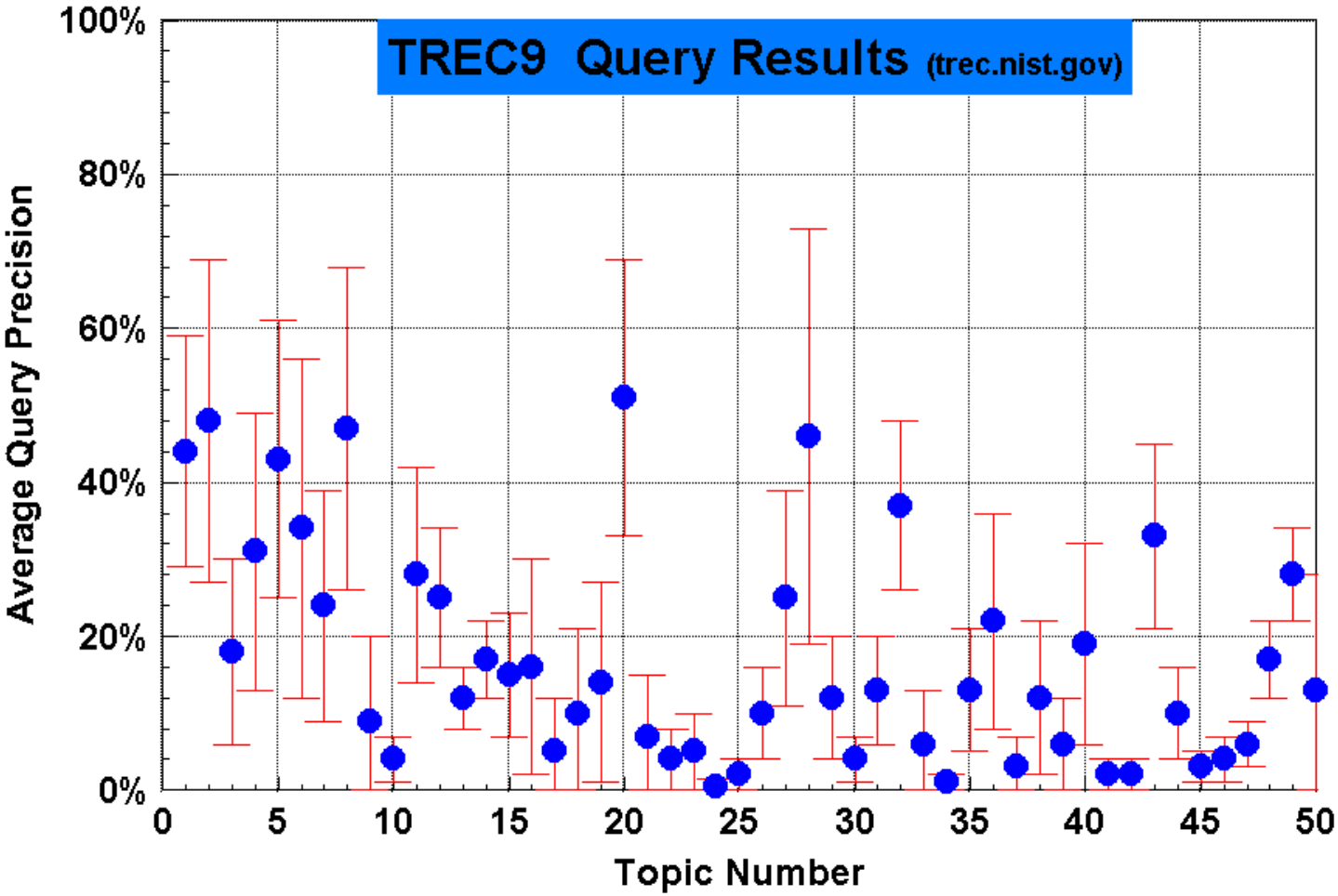
Potential KM System Solutions

- Use people: librarians, intelligence analysts
  - Very effective but costly and need time
  - Consistency and interoperability
- Limit content: small validated set
  - Hard to maintain currency
  - Large effort to collect, review, organize
  - Who decides what is authoritative?

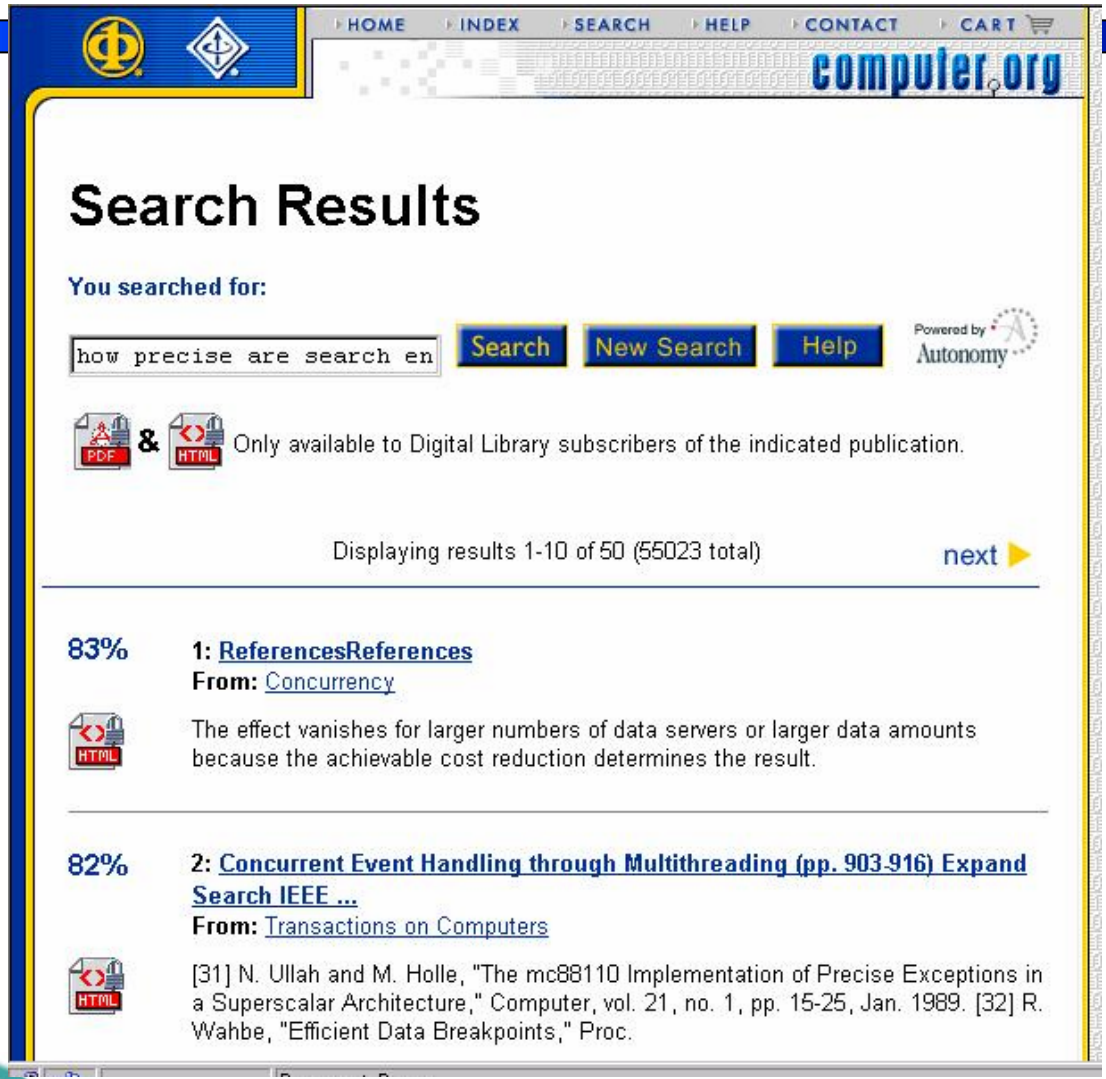
Potential KM System Solutions

- Technology: search (unstructured text)
  - Low maintenance effort but low precision
  - Depends on scaling and information density
- Use categories
  - creates smaller sets and introduces structure
    - ancient technique: Pinakes' Callimachus
  - **TAXONOMY defines topics and their relationships**
    - **improves user and technology effectiveness**

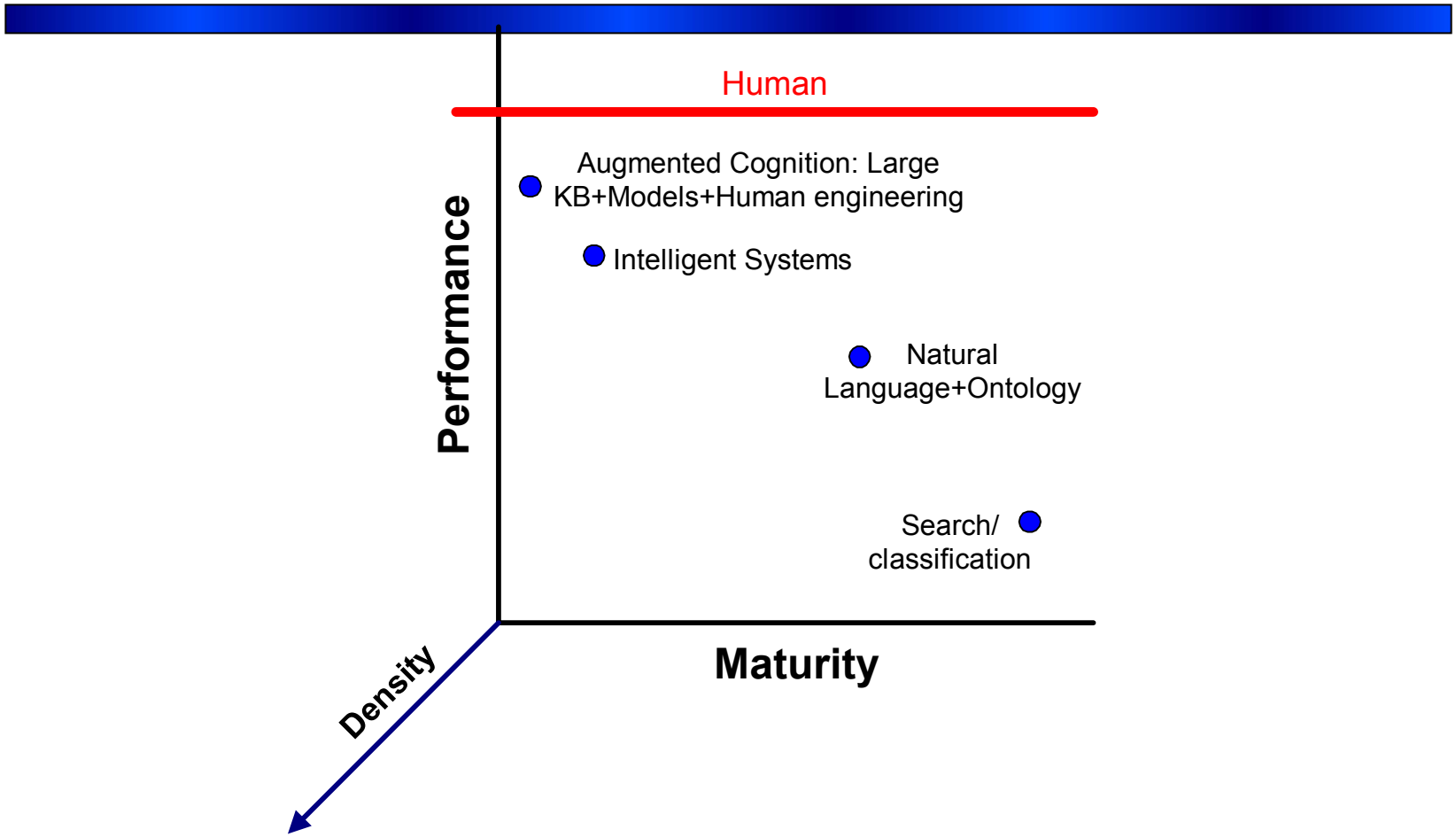
Search Engines: State-of-the-Art



Tool Performance



Technology Performance Capabilities



Organizing Information

- Ontologies and taxonomies
  - concepts and descriptions
- Develop enterprise architecture for organizational taxonomies
  - Every workgroup naturally develops its own most efficient schema
- People mentally organize in multiple ways based on task and interest

Taxonomy Definition: APQC

- A classification scheme for the knowledge accessible through a given system or interface (ultimately multi-dimensional)
- Facilitates effective retrieval, capturing, and recognition of content that is important to target users
- A taxonomy typically includes:
  - A navigable hierarchy of concepts and terms
  - Information "tags" that further identify and categorize content elements
- Links from the taxonomy lead to resources (e.g., people, documents, and events)
  - May or may not also include a thesaurus

Datawarehouse MetaData

Examples of Technical Meta Data
User report and query access patterns, frequency, and execution time
System audit controls and balancing information
The system of record feeding the data warehouse
Identification of source system fields
Mappings and transformations from the system of record to the data warehouse

Examples of Business Meta Data
The structure of data as known to the business analyst
Common access routines for information in the warehouse/mart
Table names and definitions in business terms
Attribute names and definitions in business terms

David Marco, Enterprise Warehousing Solutions, Inc., January 1999

Taxonomy Complexity

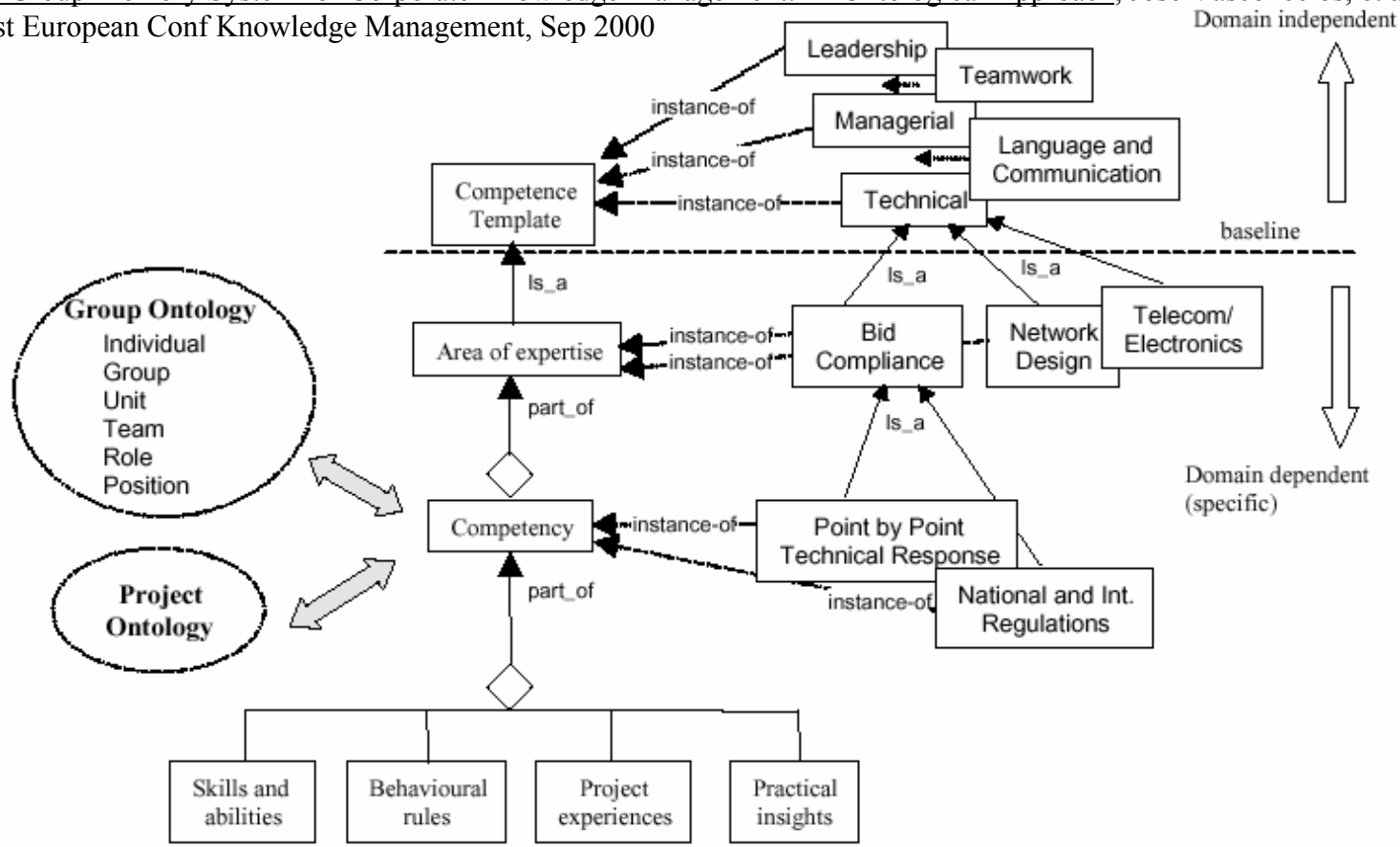
80. INTERDISCIPLINARY PHYSICS AND RELATED AREAS OF SCIENCE AND TECHNOLOGY

81. Materials science

81.05.Ft	Specific materials: fabrication, treatment, testing and analysis
∀∀∀∀	*Superconducting materials, see 74.70 and 74.72*
∀∀∀∀	*Magnetic materials, see 75.50*
∀∀∀∀	*Optical materials, see 42.70*
∀∀∀∀	*Dielectric, piezoelectric, and ferroelectric materials, see 77.80*
∀∀∀∀	*Colloids, gels, and emulsions, see 82.70.D, G, K respectively*
∀∀∀∀	*Biological materials, see 87.14*
81.05.Bx	Metals, semimetals, and alloys
81.05.Cy	Elemental semiconductors
81.05.Dz	II–VI semiconductors
81.05.Ea	III–V semiconductors
81.05.Gc	Amorphous semiconductors
81.05.Hd	Other semiconductors
81.05.Je	Ceramics and refractories (including borides, carbides, hydrides, nitrides, oxides, and silicides)
81.05.Kf	Glasses (including metallic glasses)
81.05.Lg	Polymers and plastics; rubber; synthetic and natural fibers; organometallic and organic materials
81.05.Mh	Cermets, ceramic and refractory composites
81.05.Ni	Dispersion-, fiber-, and platelet-reinforced metal-based composites

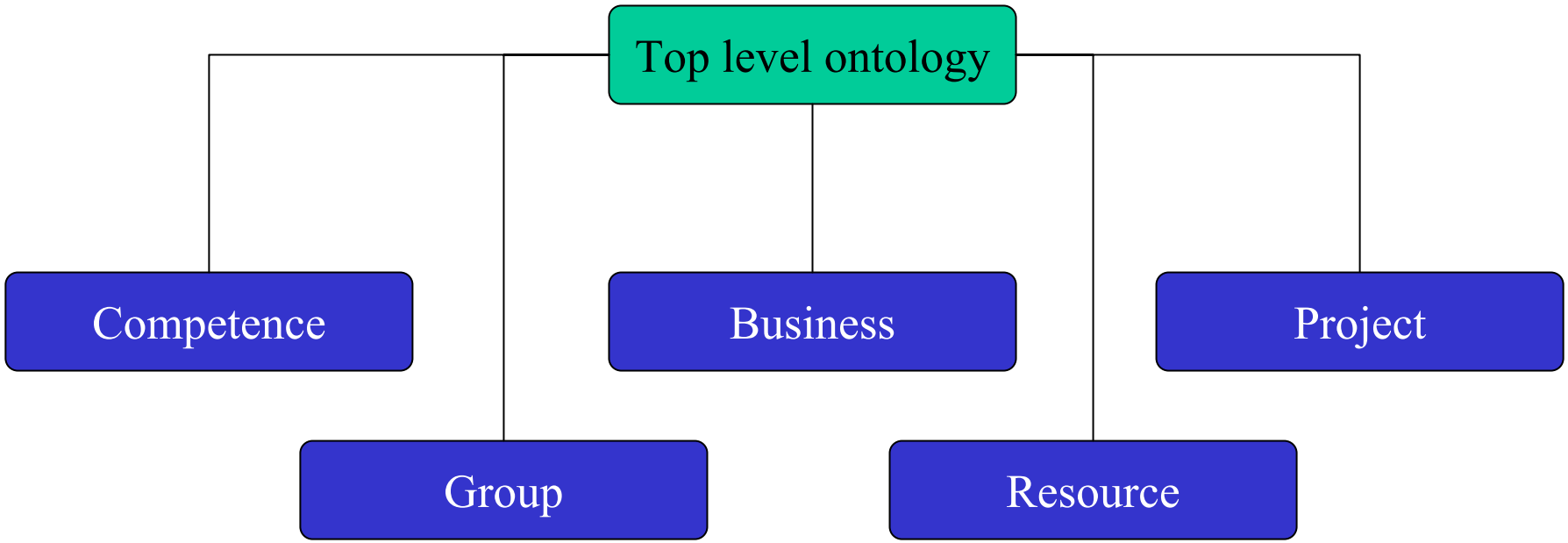
Organization Requires Context

A Group Memory System for Corporate Knowledge Management:An Ontological Approach, José Vasconcelos, et al, Proc 1st European Conf Knowledge Management, Sep 2000

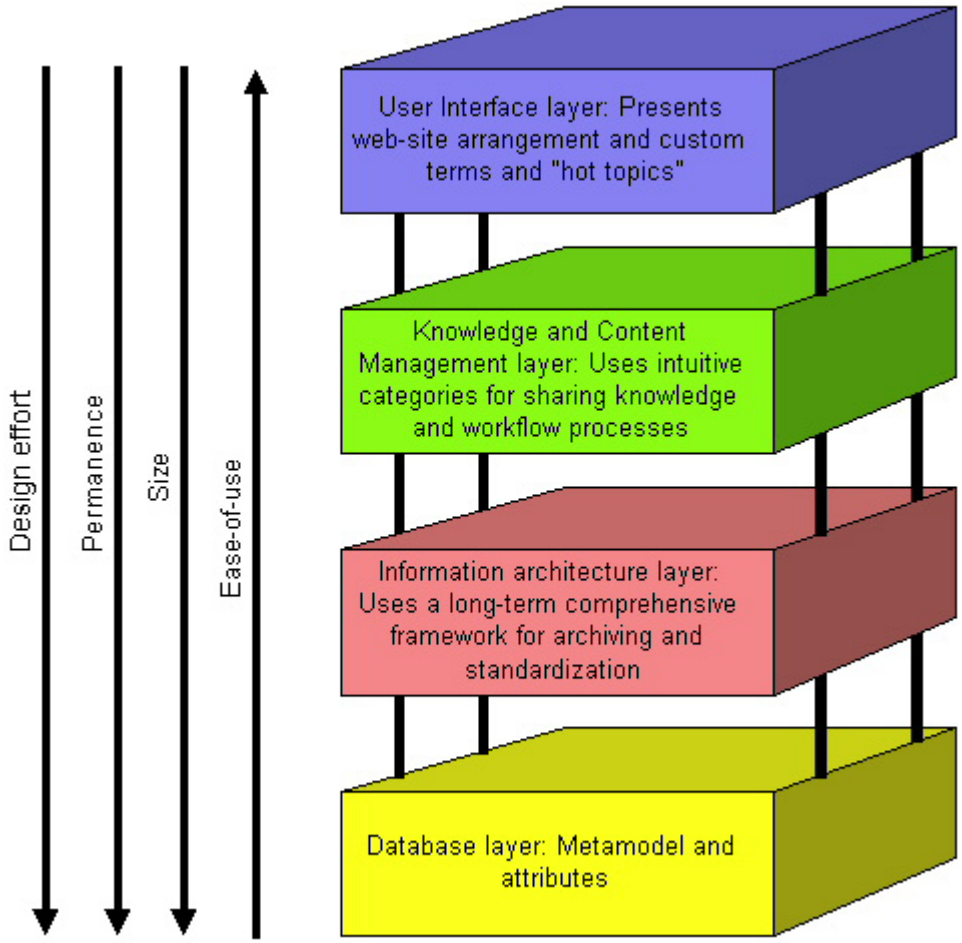


Different Perspectives

A Group Memory System for Corporate Knowledge Management:An Ontological Approach, José Vasconcelos, et al, September 2000,



Types of Taxonomies

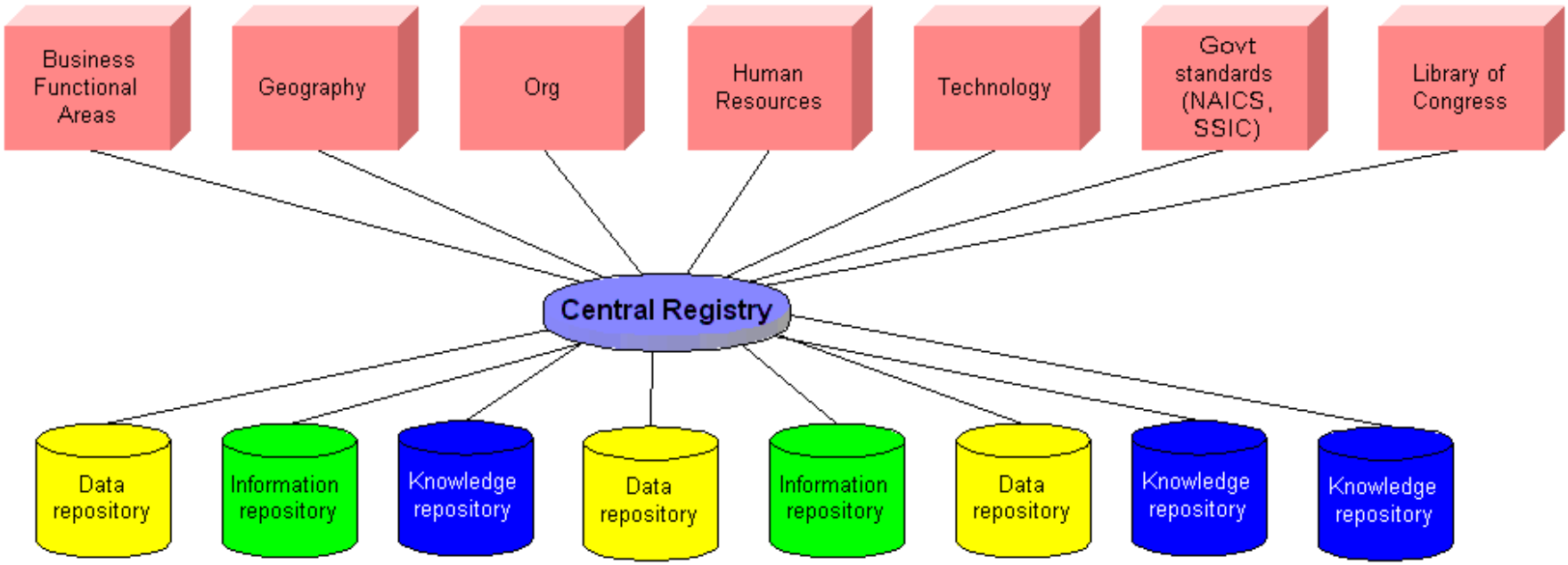


KM Taxonomy

What is a KM taxonomy?

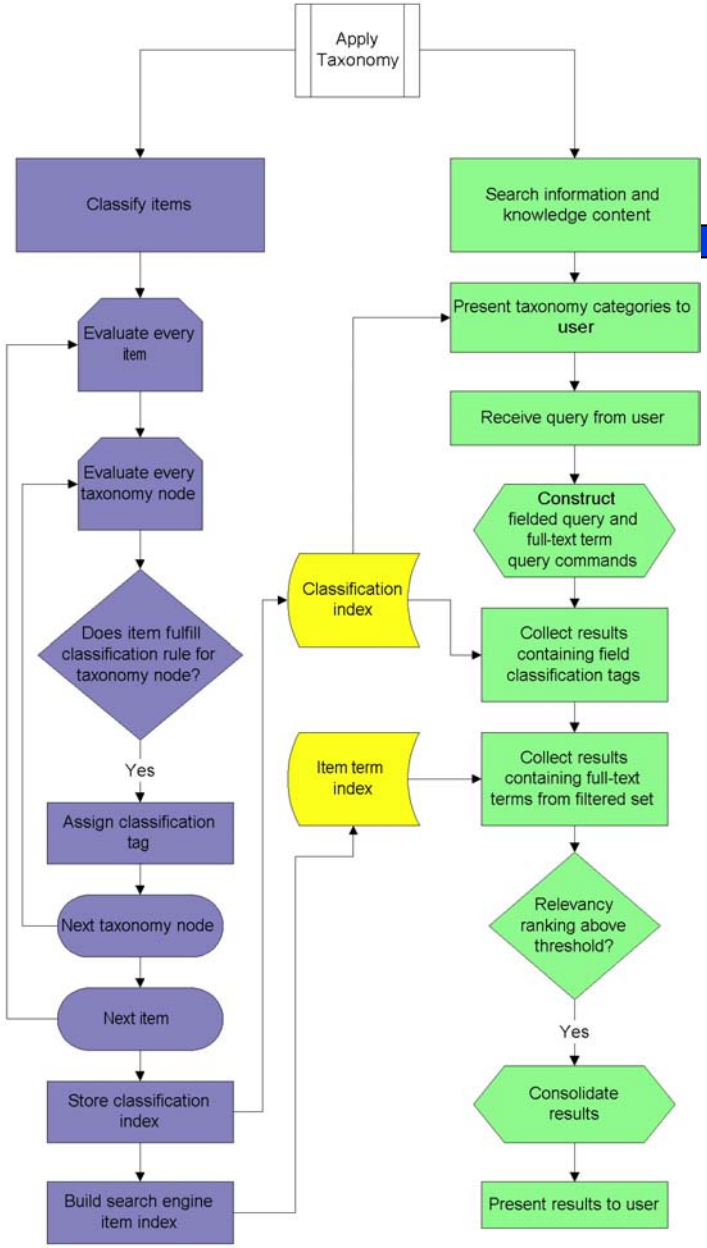
A taxonomy is a structured set of names and descriptions used to organize sources in a consistent way. A typical taxonomy uses a logical arrangement but doesn't account for users' particular decision-making and action-taking needs. A KM taxonomy focuses on enabling efficient and interoperable retrieval and sharing of data, information, and knowledge across the enterprise by building in natural workflow and knowledge needs in an intuitive structure.

Multiple Domains



Applying Taxonomy

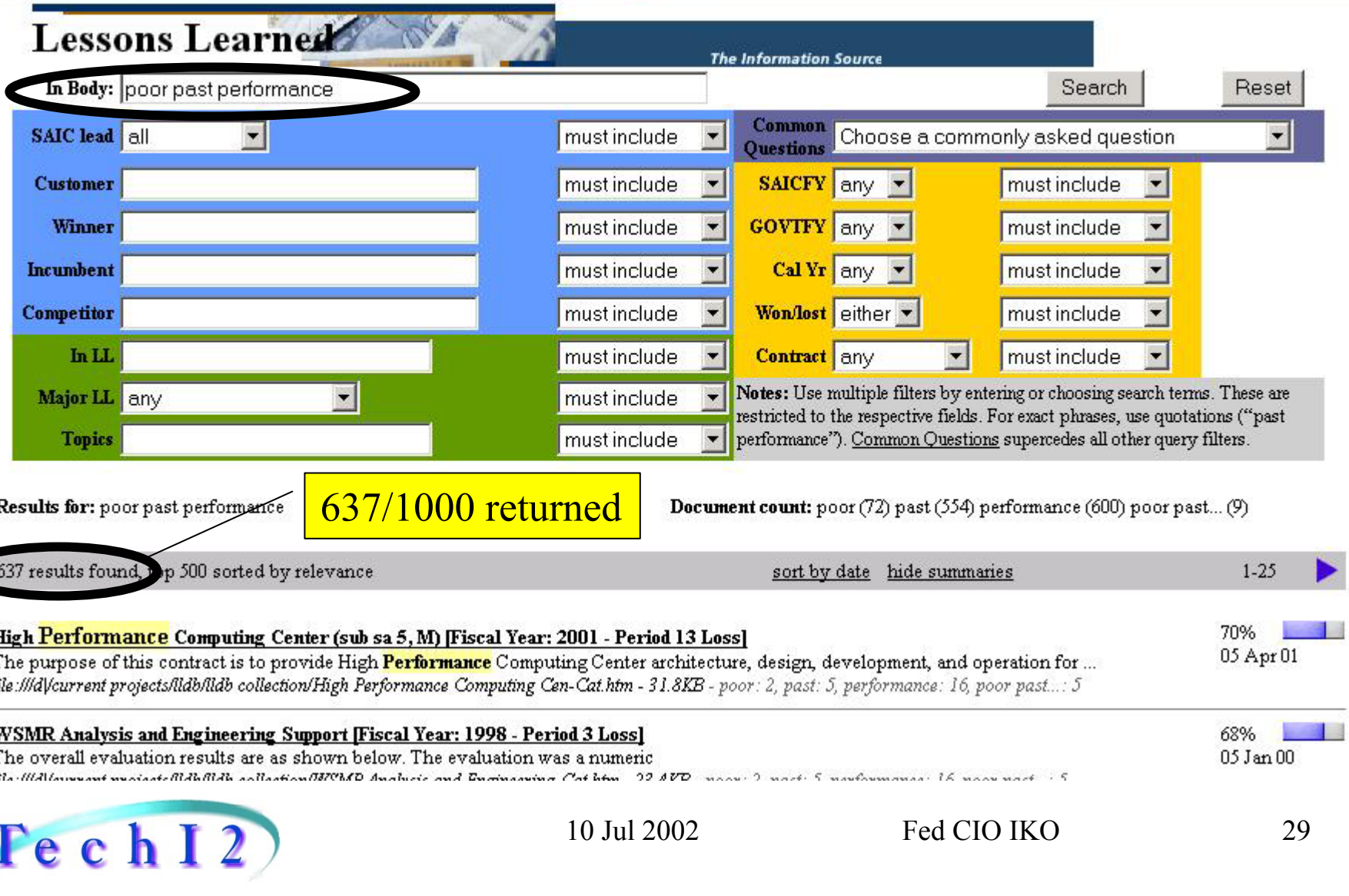
- Must associate taxonomy nodes with every document/record/multimedia/…
- How accurate is the assignment?
- Taxonomy use: What/why/who/
  - work processes and applications
- Where: HTML header, metadata repository
- How:
  - Fielded search engine query
  - Middleware


Apply Taxonomy
Classify items
Search information and knowledge content
Evaluate every item
Evaluate every taxonomy node
Does item fulfill classification rule for taxonomy node?
Yes
Assign classification tag
Next taxonomy node
Next item
Store classification index
Build search engine item index
Classification index
Item term index
Present taxonomy categories to user
Receive query from user
Construct fielded query and full-text term query commands
Collect results containing field classification tags
Collect results containing full-text terms from filtered set
Relevancy ranking above threshold?
Yes
Consolidate results
Present results to user

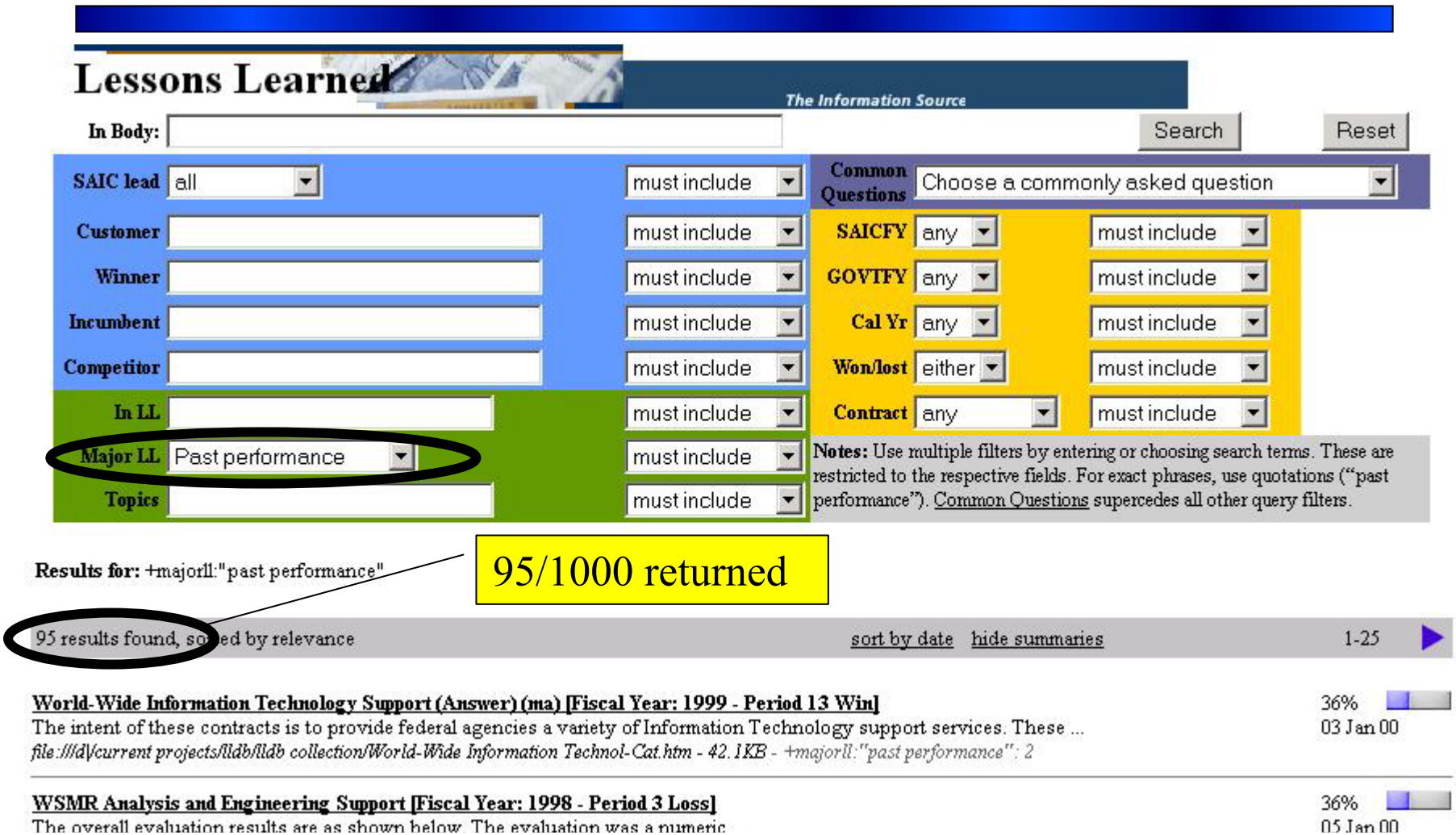
Example


Lessons Learned
The Information Source
In Body:
Search
Reset
SAIC lead
all
must include
Common Questions
Choose a commonly asked question
Customer
must include
SAICFY
any
must include
Winner
must include
GOVTFY
any
must include
Incumbent
must include
Cal Yr
any
must include
Competitor
must include
Won/lost
either
must include
In LL
must include
Contract
any
must include
Major LL
any
must include
Topics
must include
Notes: Use multiple filters by entering or choosing search terms. These are restricted to the respective fields. For exact phrases, use quotations ("past performance"). Common Questions supercedes all other query filters.

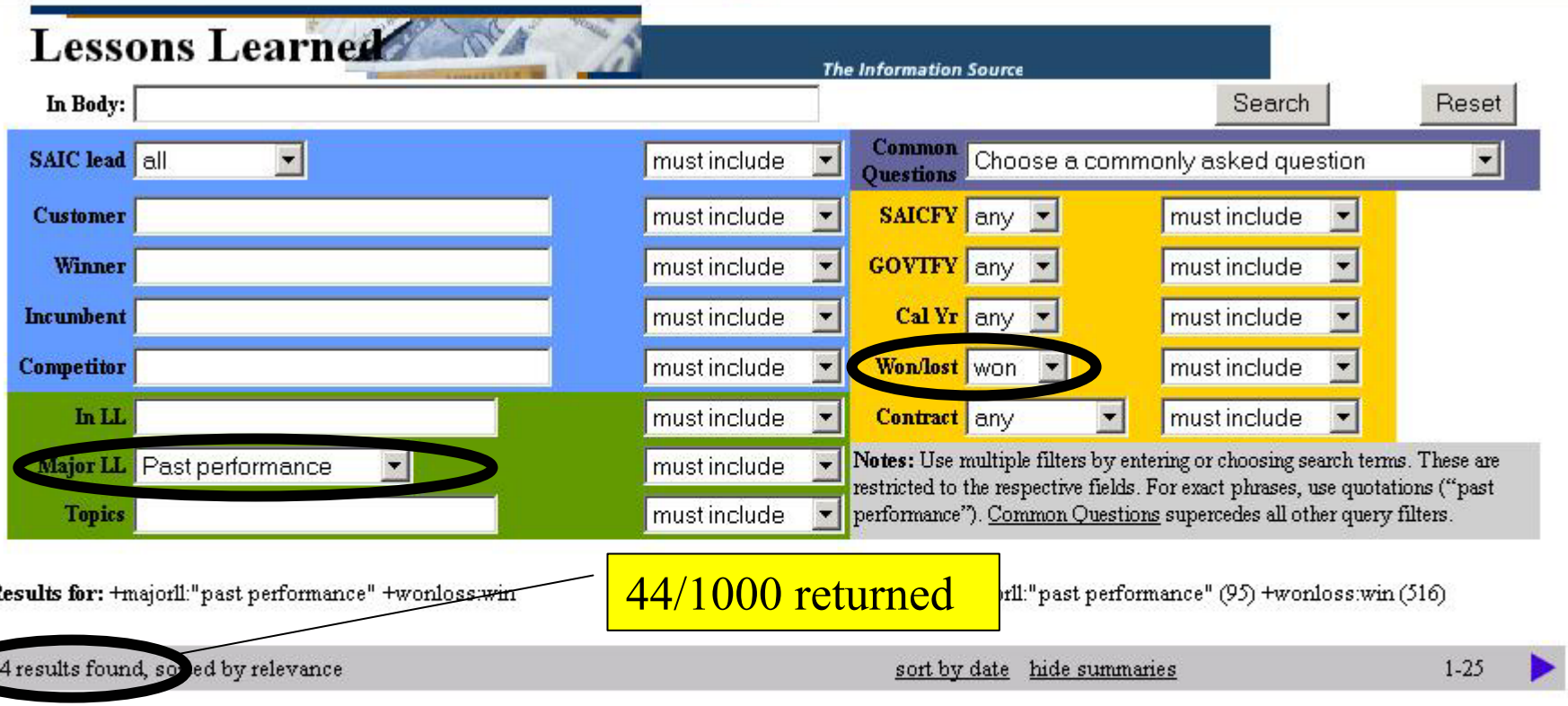
Demonstration: Typical Search



Demonstration: 1 KM Category

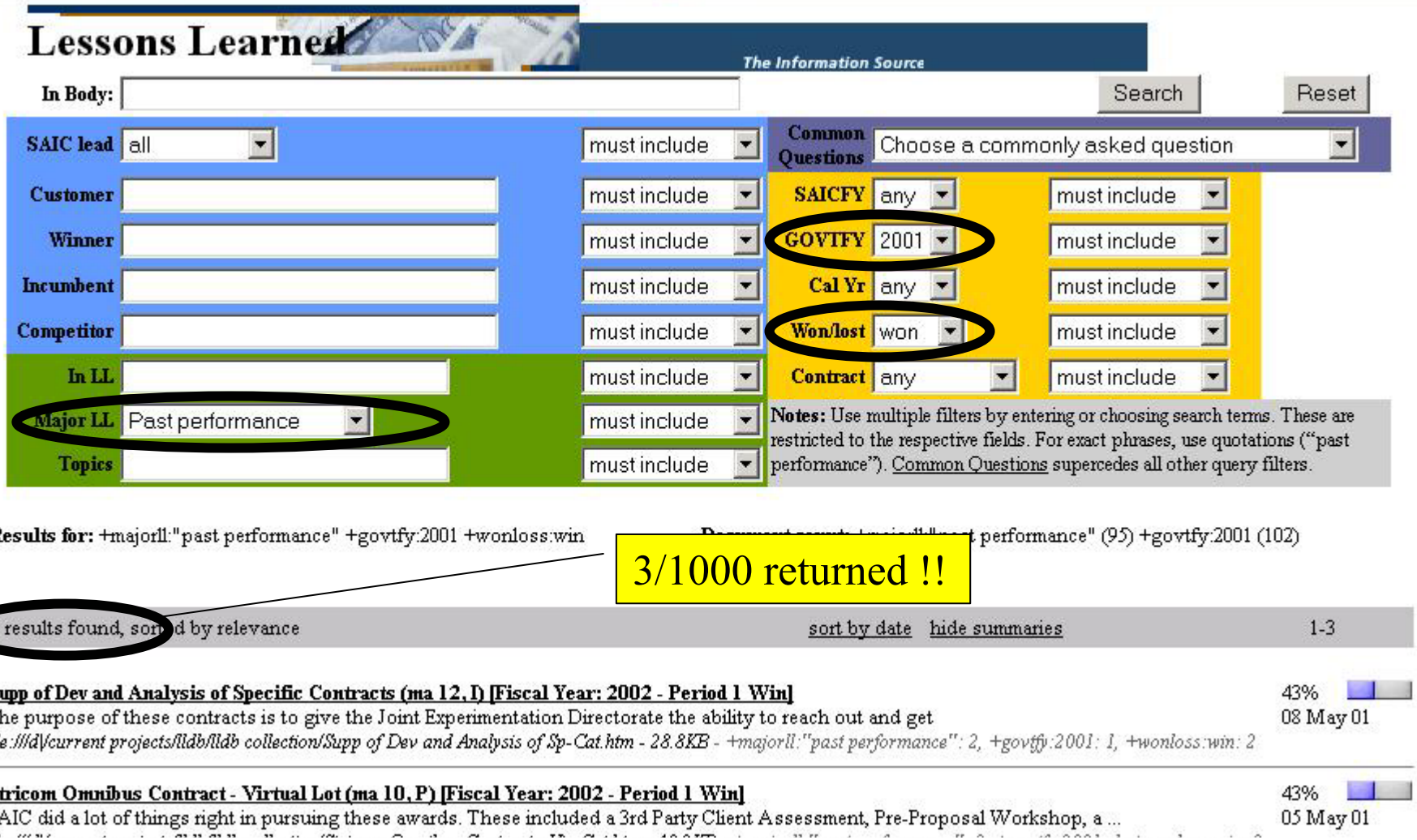


Demonstration: 2 Intuitive Categories



44/1000 returned

Demonstration: Simply Add More



Conclusions

- Automated search/categorization is still not precise enough
- Taxonomies allow large information quantity to be separated into more manageable chunks
- There are different types of taxonomies
- Effectiveness and usability require KM Taxonomy of intuitive categories